

Identifying wind power ramp causes from multivariate datasets: a methodological proposal and its application to reanalysis data

Cristobal Gallego-Castillo¹ Elena Garcia-Bustamante², Alvaro Cuerva¹, Jorge Navarro³

¹DAVE, Universidad Politécnica de Madrid, Pza Cardenal Cisneros 3, 28040, Madrid, Spain

²Departamento de Física, Universidad de Murcia, Campus Espinardo, s/n, 30100, Murcia, Spain

³Unidad de Energía Eólica, CIEMAT, Avenida Complutense, 40, 28040, Madrid, Spain

Abstract: Forecasting abrupt variations in wind power generation (the so-called ramps) helps achieve large scale wind power integration. One of the main issues to be confronted when addressing wind power ramp forecasting is the way in which relevant information is identified from large datasets to optimally feed forecasting models. To this end, an innovative methodology oriented to systematically relate multivariate datasets to ramp events is presented. The methodology comprises two stages: the identification of relevant features in the data and the assessment of the dependence between these features and ramp occurrence. As a test case, the proposed methodology was employed to explore the relationships between atmospheric dynamics at the global/synoptic scales and ramp events experienced in two wind farms located in Spain. The achieved results suggested different connection degrees between these atmospheric scales and ramp occurrence. For one of the wind farms, it was found that ramp events could be partly explained from regional circulations and zonal pressure gradients. To perform a comprehensive analysis of ramp underlying causes, the proposed methodology could be applied to datasets related to other stages of the wind-to-power conversion chain.

1 Introduction

One of the main drawbacks of wind energy is that it exhibits intermittent generation greatly depending on environmental conditions. Wind power forecasting has proven to be an effective tool for facilitating wind power integration from both the technical and the economical perspective [1, 2]. Research on this field is mainly concerned with the modelling of wind power dynamics, these resulting from the interaction between several processes taking place in the wind-to-power conversion chain. Fig. 1 depicts this chain, reflecting the meteorological stage (the local wind as a result of atmospheric processes occurring at different temporal/spatial scales) and the technological stage (the conversion of local wind into electricity by the wind turbines).

An issue of growing interest over the last few years is that of wind power ramp forecasting [3]. Ramp events represent a particular case of wind power dynamics characterised by a large and rapid variation (1–4 h) observed in the power output of a wind farm or portfolio. Due to its very nature, ramps may derive in large forecasting errors, with the negative consequences from the economic and the technical perspective [4]. Hence wind power forecasting traditionally focuses on minimising the overall error committed in large time periods, the impact of such extreme situations may be occasionally undervalued by forecasters. Consequently, wind power ramp forecasting arises as a need for improving forecasts during these events.

As a particular case of wind power dynamics, ramp events are conditioned to processes occurring at any stage of the wind-to-power conversion chain. Indeed, ramps can be motivated by different meteorological phenomena (from large-scale fronts to local phenomena such as thunderstorms and thermally-driven flows), but also by specific conditions of the wind farm operation (rotor yaw misalignment, wind turbine shutdown, rotor-wake interactions etc.). Monitoring/modelling the stages involved in the wind-to-power conversion chain generates massive multivariate datasets that can be employed to explore ramp underlying causes, and eventually improve ramp forecasting. For example, outputs

generated with numerical weather prediction (NWP) models represent with different levels of realism the atmosphere dynamics that ultimately are responsible for the climate variability at different spatial scales: general circulation models (GCMs), due to their coarser horizontal resolution, provide estimates of the climate system that are physically coherent at global/hemispheric and continental spatial scales, while regional circulation models (RCMs), with increased spatial resolution with respect to GCMs, provide estimations of the regional climate variability by resolving processes at smaller spatial scales over specific areas. In addition, observational data gathered at weather stations characterise local weather conditions, and data gathered by the SCADA (supervisory control and data acquisition) system embody information about the wind farm state. However, all these datasets are likely to contain large amounts of irrelevant information for ramp explanation and the question of how to identify potential explanatory variables oriented to ramp forecasting is nowadays relatively unexplored.

Some authors have contributed to a better understanding of wind power ramp underlying causes by analysing qualitatively some of the aforementioned stages, specially atmospheric phenomena. In this vein, an interesting classification of the main meteorological processes that may derive in ramp events and their predictability was provided in [5]. Cutler *et al.* [6] reported one of the first attempts to categorise wind power ramps for the case of a single wind farm, located in North-West coast of Tasmania. To this end, NWP outputs provided by a RCM were employed. Another interesting case study is the pilot project in the Alberta Electric System Operator (AESO) area (Canada), a relatively flat region situated close to a large mountain range. In this case, orographic effects were deemed to be determinant in explaining ramp events [7]. More recently, Bossavy *et al.* [8] applied NWP ensembles to analyse the temporal uncertainty of ramp occurrence. To this end, the case of a single wind farm located in complex terrain in France was considered. Based on meteorological outputs provided by the Ensemble Prediction System of the European Centre for Medium-Range Weather Forecasts (ECMWF), the authors obtained the probabilities of ramp occurrence for different sets of time intervals.

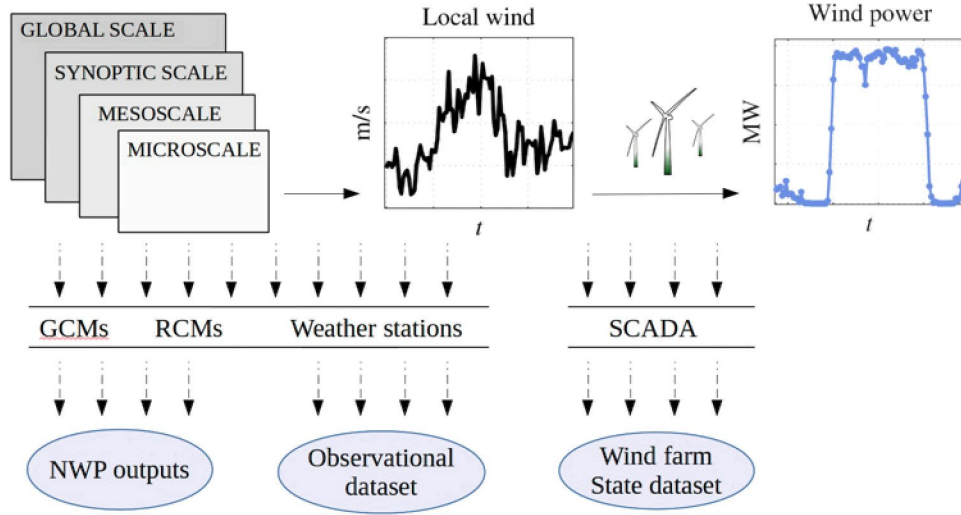


Fig. 1 Wind-to-power conversion process and datasets obtained from monitoring/modelling the different stages

Focusing on the analysis of meteorological records, Kamath [9] analysed the relation between daily averaged weather measurements and ramp events in the Tehachapi Pass and the Columbia Basin, two regions with important wind penetration levels. The author concluded that wind measurements (in particular, average wind speed, speed gusts and wind direction) were of main interest in detecting days with ramp occurrence in both regions. Finally, a few works have analysed ramp event causes and predictability based merely on power data gathered by the SCADA system [10, 11].

Due to ramp event underlying causes are likely to depend strongly on specific-case features (local climate, portfolio size, wind turbine layout etc.), expertise gained from the aforementioned qualitatively analyses cannot be easily generalised to other case studies. To our knowledge, a more general approach for exploring ramp causes within the wind-to-power conversion chain is still required. With this purpose, an innovative methodology to systematically relate massive datasets to wind power ramp occurrence is here introduced. It consists of a two-staged approach: first, principal component analysis (PCA) is employed to identify relevant features of the raw dataset. Second, the notion of mutual information (MI) is employed to identify those features showing maximum dependence with ramp occurrence. The proposed methodology can be applied to any dataset originated from the wind-to-power conversion process. As a particular case, the contribution of the global/synoptic scales to ramp occurrence at the wind farm level is here considered.

The remaining of this paper is organised as follows. Section 2 provides a general description of the proposed methodology. The employed datasets are detailed in Section 3. The obtained results are presented and discussed in Section 4. The paper ends in Section 5 with the main conclusions of the work.

2 Methodology

As mentioned above, ramps events represent a particular case of wind power dynamics characterised by an abrupt change of the power output. These events can be motivated by processes occurring at any stage of the wind-to-power conversion chain. These underlying processes are assumed to be properly captured in large datasets generated by means of monitoring/modelling techniques (see Fig. 1). The proposed methodology aims to systematically search for relationships between a certain massive dataset, denoted by X hereafter, and the wind power ramp events. Specifically, X is a matrix in $\mathbb{R}^{N \times J}$, gathering J time series organised column-wise, in such a way that $x_{t,j}$ with $1 \leq t \leq N$ and $1 \leq j \leq J$ represents the t th value of the j th time series.

Mathematically

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,J} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N,1} & x_{N,2} & \cdots & x_{N,J} \end{bmatrix}. \quad (1)$$

It is assumed that ramp events observed within a wind power time series can be characterised in some way by a one-dimensional (1D) time series, here denoted by $\mathbf{r} = [r_1, r_2, \dots, r_N]^T$. There exist several approaches to define r_t ($1 \leq t \leq N$), in the form of either a binary or a continuous-valued function. A detailed review can be found in [12], where the authors discussed some of the drawbacks related to the use of binary definitions and introduce the so-called ramp function as a means to characterise wind power ramp events in a continuous range. The idea behind this function is to evaluate the ramp intensity at each time step by considering the power gradient observed within several time scales. Mathematically, the ramp function at time t is given by

$$r_t(\lambda_N) = \sum_{\lambda=2}^{\lambda_N} W_H^{t,\lambda}, \quad (2)$$

where λ_N is a parameter that customises the range of time scales in which ramps are to be captured and $W_H^{t,\lambda}$ is the wavelet transform of the wind power time series based on the Haar function, given by

$$W_H^{t,\lambda} = \begin{cases} \frac{1}{\sqrt{\lambda}} \cdot \left(\sum_{j=1}^{\lambda/2} p_{t+j-1} - \sum_{j=1}^{\lambda/2} p_{t-j} \right), & \text{if } \lambda \text{ is even} \\ \frac{1}{\sqrt{\lambda}} \cdot \left(\sum_{j=1}^{(\lambda-1)/2} p_{t+j} - \sum_{j=1}^{(\lambda-1)/2} p_{t-j} \right), & \text{if } \lambda \text{ is odd} \end{cases} \quad (3)$$

where p_t is the power record at time t . For further details on the ramp function, readers are referred to [12]. From here on, r_t is assumed to be the ramp function customised to capture ramp events of durations between 1 and 4 h ($\lambda_N = 5$) and re-scaled between -1 and 1 , for convenience. Fig. 2 shows the ramp function of the wind power time series of wind farm A (see Section 3 for details) during a 4 days period.

Our objective here is to explore the extent to which processes captured in the dataset underlie ramp events, that is, if there exists some kind of link between X and \mathbf{r} . With this purpose, a two-staged methodology is introduced in the following subsections.

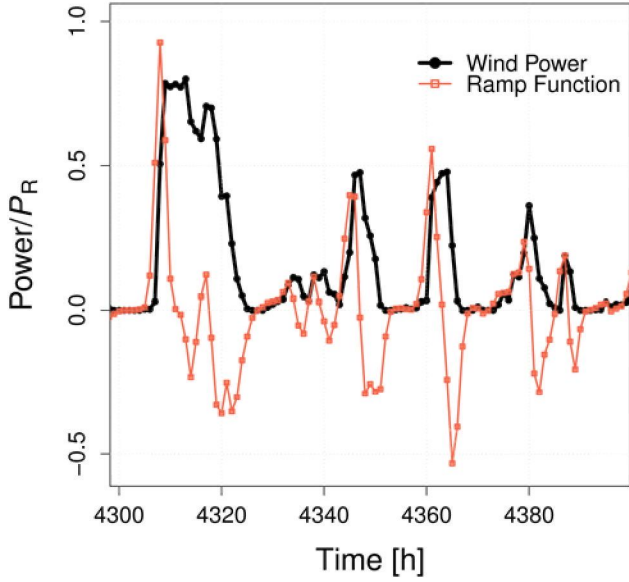


Fig. 2 Ramp function at wind farm A computed for $\lambda_N = 5$

2.1 Identifying features in the original dataset

PCA is a statistical technique oriented to reveal the underlying structure of the data through data compression [13]. The idea behind PCA is that raw data do not represent an efficient way to store information if the involved variables are correlated. Thus, through an appropriate linear transformation, it is possible to obtain a new set of variables which are mutually uncorrelated, in such a way that a reduced number of these variables (those related to the main modes of variability) accounts for the maximum possible variance of the original dataset.

In practice, by applying PCA to X , J modes (also referred to as empirical orthogonal functions, EOFs) together with J principal components (PCs) are obtained. The EOFs are vectors in \mathbb{R}^J . PCs are time series of length N , showing how the original data are organised into modes across time. In this work, \mathbf{eof}_j denotes the j th EOF, \mathbf{pc}_j is the associated PC and λ_j represents the percentage of the variability explained by the related mode. In addition, modes are to be organised in decreasing order of the explained variance. By doing this, first modes can be thought of as the main features of the original dataset as they show the way in which data combine to jointly exhibit maximum variance. It holds that

$$[\mathbf{pc}_1, \mathbf{pc}_2, \dots, \mathbf{pc}_J] = X \cdot [\mathbf{eof}_1, \mathbf{eof}_2, \dots, \mathbf{eof}_J] \quad (4)$$

and $\lambda_i \geq \lambda_j$ for any $1 \leq i < j \leq J$.

A final remark needs to be done: PCA assumes that matrix X is composed by zero-mean variables. In the remaining of this paper, the use of standardised variables is assumed (i.e. each variable has been rescaled to zero mean and unit variance).

2.2 Assessing the relationship between data features and ramp events

The second part of the methodology consists in assessing the dependence between the aforementioned main features of the dataset and ramp occurrence. To this end, the notion of MI is employed. MI is thought of as a reliable tool for assessing non-linear dependence between random variables without any request about the model of dependency [14, 15]. MI is an outcome of the information theory [16]. Within this context, the notion of *entropy* of a random variable Y , denoted by $H(Y)$, is used as a measurement of the uncertainty related to Y , while *conditional entropy*, $H(Y|Z)$, represents the uncertainty of Y given the outcome of another random variable Z . By combining these two concepts,

the MI between Y and Z , defined as

$$MI(Y, Z) = H(Y) - H(Y|Z), \quad (5)$$

establishes how much the outcome of Z tells us about the outcome of Y , which enables a measurement of the dependence between both variables. In practice, for the case of continuous random variables, $MI(Y, Z)$ is defined as follows

$$MI(Y, Z) = - \iint f_{YZ}(y, z) \log \left[\frac{f_{YZ}(y, z)}{f_Y(y)f_Z(z)} \right] dy dz, \quad (6)$$

where $f_Y(y)$ and $f_Z(z)$ are the probability density functions (pdf) of Y and Z , and $f_{YZ}(y, z)$ is the joint pdf of the mentioned random variables.

In view of these aspects, in order to estimate the dependence between the j th mode obtained through PCA, \mathbf{eof}_j , and ramp occurrence, we assume that the time series \mathbf{pc}_j and \mathbf{r} represent an ordered sequence of N realisations of two random variables, PC_j and R respectively, with some kind of (non-linear) dependence between them. Then, through (6) and density estimation techniques, it is possible to compute $MI(PC_j, R)$.

Finally, we point out that the MI is likely to be overestimated when it is computed from a finite number of samples [17]. In this work, the associated bias will be assessed following the expressions detailed in [15].

3 Description of the datasets

The remaining of this paper deals with a particular application of the proposed methodology. With this aim, two sets of data were considered. One refers to data gathered by the SCADA system for two wind farms located in Spain. The second dataset originates from the reanalysis performed by the National Centers for Environmental Prediction (NCEP).

3.1 Wind farm data

Two wind farms located in the Iberian Peninsula have been considered in this work. They are referred to as wind farms A and B (see Fig. 3). The rated power of the wind farms are 24.5 and 36 MW, respectively. Measurements of wind power output averaged on an hourly basis were available for each wind farm. In

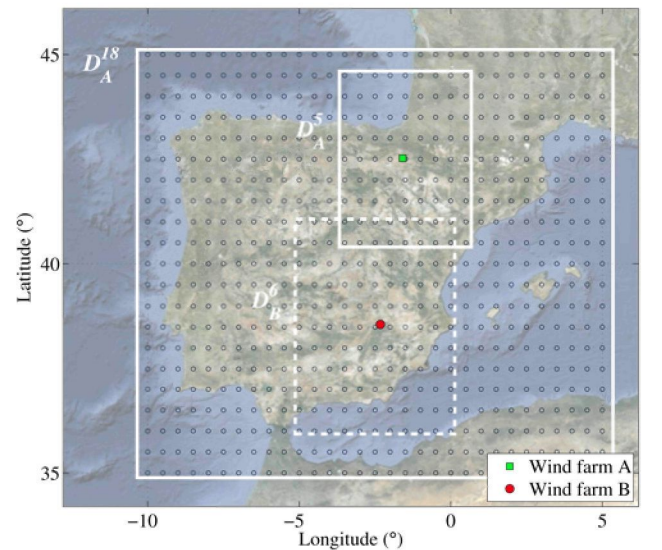


Fig. 3 Wind farm locations. Circles show the nodes of the grid where meteorological data from the NCEP reanalysis were available for this work. Spatial domains D_A^5 , D_A^{18} and D_B^6 are highlighted (see text for details)

the case of wind farm A, the period considered ranges from 1 November 2007 to 16 September 2008. For the case of wind farm B, wind power data were obtained between 1 March 2007 and 23 March 2008.

A problem concerning wind power modelling is the identification of wind power dynamics due to abnormal operation (breakdown, maintenance, untwisting sequence of power cables or transmission system operator (TSO) stop order). To remove these data, such periods were identified in the base of an additional dataset reporting the status of every wind turbine on a 10-minute basis. As a result, the percentage of valid wind power data was found to be 93.43% for wind farm A and 92.51% for wind farm B. It is noted that some ramp events observed in the original wind power time series were actually due to abnormal operation. More details on the data-cleaning process can be found in [18].

3.2 Meteorological data

The meteorological dataset employed in this work originates from the reanalysis performed by NCEP [19]. This reanalysis has two main advantages as compared with other atmospheric reanalyses (see [20] for a list of current reanalyses). First, the horizontal resolution of 0.5° represents the highest of the reanalysis datasets available at the time of the elaboration of this work. Second, the time series of the meteorological variables are provided with a 1 h resolution, matching with the wind power time series resolution. This issue represents a strong motivation for selecting the mentioned reanalysis dataset, as one of the goals of the study is to seek potential relationships between atmospheric variables and local wind power events (ramps events) characterised by a duration of around 1–4 h.

The meteorological dataset consists of 5859 hourly time series, resulting from considering a grid with 21×31 nodes and nine meteorological variables per node. The grid covers the Iberian Peninsula (latitudes from 35°N to 45°N and longitudes from 10°W to 5°E , see Fig. 3) with the aforementioned resolution of 0.5° . The nine meteorological variables, detailed in Table 1, are the two components of the horizontal wind speed at three vertical levels (10 m above sea level, 850 and 500 hPa), mean sea-level pressure (MSLP) and geopotential height at two vertical levels (850 and 500 hPa). These meteorological variables are usually thought of as natural candidates for explaining the wind variability at the surface [21, 22]. Similar meteorological variables have been considered in previous works to explain wind power dynamics (see [9, 23, 24], among others).

4 Application to GCM outputs

In this section, the methodology introduced in Section 2 is utilised to identify information in the global/synoptic scales relevant for explaining power ramp occurrence at the wind farm level. To this end, datasets described in Section 3 are employed.

4.1 Meteorological feature identification

PCA is to be applied to a variety of subsets of the meteorological dataset. These subsets arise from considering different combinations of meteorological variables and spatial domains, as follows:

(a) *Variable selection*: First, the meteorological variables considered in the PCA are selected. Eight cases are defined:

- (a.1) UV_{10} : Wind field at 10 m a.s.l. Two meteorological variables are involved: U_{10} and V_{10} .
- (a.2) UV_{850} : Wind field at 850 hPa (U_{850} and V_{850}).
- (a.3) UV_{500} : Wind field at 500 hPa (U_{500} and V_{500}).
- (a.4) UV_s : Wind field at the three levels above mentioned (U_{10} , V_{10} , U_{850} , V_{850} , U_{500} and V_{500}).
- (a.5) $MSLP$: Mean sea-level pressure.
- (a.6) Z_{850} : Geopotential height at 850 hPa.
- (a.7) Z_{500} : Geopotential height at 500 hPa.

Table 1 Meteorological variables from NCEP reanalysis employed in this work

Variable	Units	Description
U_{10}	(m s ⁻¹)	zonal wind component at 10 m a.s.l. (above sea level)
V_{10}	(m s ⁻¹)	meridional wind component at 10 m a.s.l.
U_{850}	(m s ⁻¹)	zonal wind component at pressure level 850 hPa ^a
V_{850}	(m s ⁻¹)	meridional wind component at pressure level 850 hPa
U_{500}	(m s ⁻¹)	zonal wind component at pressure level 500 hPa ^b
V_{500}	(m s ⁻¹)	meridional wind component at pressure level 500 hPa
$MSLP$	(hPa)	mean sea-level pressure
Z_{850}	(m)	geopotential height at 850 hPa
Z_{500}	(m)	geopotential height at 500 hPa

^a850 hPa corresponds to a height of about 1500 m a.s.l.

^b500 hPa corresponds to a height of about 5500 m a.s.l.

(a.8) ΔZ : Geopotential thickness of the layer between pressure levels 850 and 500 hPa, given by $\Delta Z = Z_{500} - Z_{850}$. This case is considered because the mentioned thickness is related to horizontal temperature gradients, which cause vertical shear of the geostrophic winds (i.e. thermal winds) [25].

(b) *Domain selection*: To avoid a priori assumptions on the optimal spatial domain, a parametric analysis is proposed by considering increasing spatial domains centred in each wind farm. In particular:

(b.1) Wind farm A: 18 domains, referred to as D_A^δ , with $\delta = 1, \dots, 18$, are defined. The smaller one, D_A^1 , consists of a single node, the one closest to the wind farm; D_A^2 is obtained by including the surrounding nodes (hence it is a 3×3 domain centred in wind farm A) and so on. It is noted that, owing to the relative position between the wind farm location and the area covered by the meteorological dataset, domains larger than D_A^6 are no longer centred in the wind farm. The full size domain corresponds to D_A^{18} , with 21×31 nodes (all the nodes in Fig. 3).

(b.2) Wind farm B: 16 domains, referred to as D_B^δ , with $\delta = 1, \dots, 16$, are considered. The domains are defined in a similar fashion to the one described above. In this case, domains beyond D_B^8 are no longer centred in the wind farm. The full size domain corresponds to D_B^{16} .

Fig. 3 shows some of the domains considered for each wind farm.

(c) *Application of the PCA*: Selecting a case in (a) and a spatial domain in (b) yields a specific matrix X . After applying PCA to this matrix, the first four EOFs, PCs and percentages of the explained variability are retained for subsequent analysis.

Combining the number of cases given in (a) with the spatial domains described in (b) yields a total of 8×18 PCAs for the case of wind farm A and 8×16 PCAs for the case of wind farm B. As the large amount of outputs obtained from these analyses, in the following only a reduced number of cases are described for illustrative purposes (a deeper analysis can be found in [18]).

Let us consider the case for UV_{10} (a.1) in the domain D_A^5 (a grid centred in wind farm A with 9×9 nodes). Matrix X is given by (see (7))

where the first upper-index is a time index (N is the length of the time series) and the second upper-index refers to the node in the domain D_A^5 . According to the notation in (1), the number of columns is $J = 2 \times 9 \times 9 = 162$. Consequently, \mathbf{eof}_i is a vector in \mathbb{R}^{162} . This EOF can be consistently represented in the domain D_A^5 considering that its two first components are referred to the two first columns in X , i.e. U_{10} and V_{10} at the node (1×1) in D_A^5 , third and fourth components of \mathbf{eof}_i refer to U_{10} and V_{10} at the node (1×2) and so on. It is also noted that for scalar fields (e.g. MSLP), there is only one time series per node, hence the EOFs can be represented as contour maps in the related spatial domain.

Fig. 4 shows the first EOF obtained by applying PCA to UV_{10} and spatial domain D_A^5 (left) and spatial domain D_B^6 (right). It can be observed that both EOFs reflect a certain interaction between wind field and orography. In particular, the spatial structure of the picture on the left shows a strong pattern NW–SE. This pattern is

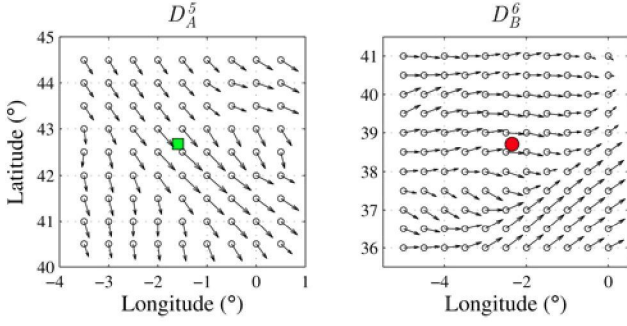


Fig. 4 First EOF obtained by applying PCA to variable UV_{10} in domain D_A^5 (left) and D_B^6 (right)

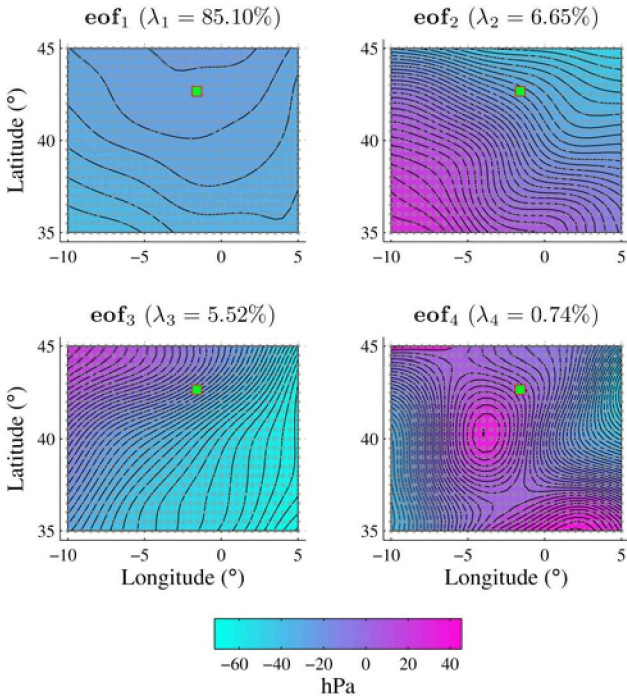


Fig. 5 First four EOFs obtained by applying PCA to MSLP in the domain D_A^{18} (for illustrative purposes, the EOFs have been re-scaled with the standard deviation). The square shows wind farm A location

likely to be related to regional wind regimes governed by the channelling effect of the Ebro valley (which is located in this area). These regimes are known as Cierzo (NW) and Bochoño (SE) [21, 22, 24]. Concerning the picture on the right, the spatial structure of the mode reveals that the Mediterranean coast plays a key role in explaining the variability of the wind field at 10 m height.

Concerning the scalar fields [cases (a.5)–(a.8)], the spatial variability of these fields was found to be noticeable lower than that of the wind fields (specially the wind at the surface). In these cases, PCA is unable to capture meaningful structures in the spatial variability of the data. Consequently, the application of the PCA yielded EOFs that were mostly related to the so-called Buell patterns [13], reflecting EOFs characterised by a homogeneous field for the first EOF followed by orthogonal dipole and tripole patterns, as a means to optimally capture the variability of the field

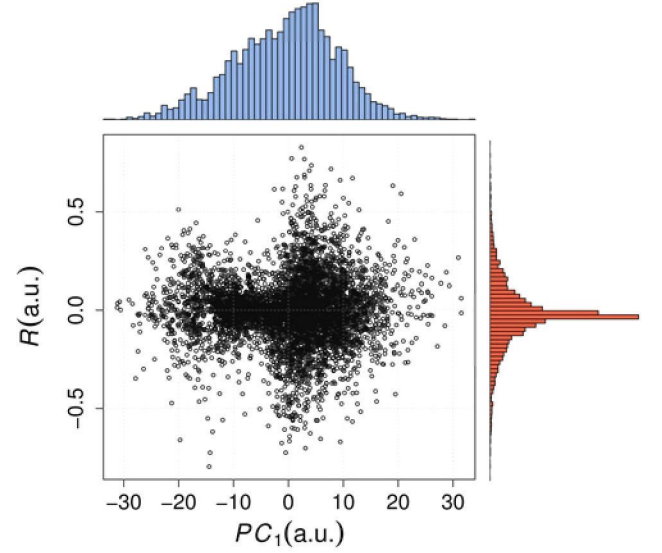


Fig. 6 pc_j and r for a particular case (see text for details). ‘a.u.’ stands for arbitrary units

given the PCA constraints. These patterns can be observed in Fig. 5, which shows the first four EOFs for the MSLP in the domain D_A^{18} . For this particular case, the first EOF can be interpreted as the average pressure level in the spatial domain (low/high pressure), which accounted for more than the 85% of the total variability, while second and third EOFs represent pressure gradients in the area (probably indicating the passage of large scale weather systems), accounting for 6.65 and 5.52% of the total variability, respectively.

4.2 Dependence between meteorological features and wind power ramp events

In this section, the dependence between meteorological features and wind power ramp occurrence at the wind farm level is explored. In particular, we shall estimate $MI(PC_j, R)$ in base of vectors \mathbf{pc}_j and \mathbf{r} . \mathbf{pc}_j represents any of the PCs obtained in Section 4.1 and \mathbf{r} represents the ramp function values of a wind power time series as described in Section 2.

Before proceeding, and for illustrative purposes, the interactions between \mathbf{pc}_j and \mathbf{r} are shown for a particular case. Fig. 6 shows the histograms and the scatter plot of the ramp function at wind farm A and the first PC obtained from applying PCA to UV_{10} at D_A^5 (the related EOF is shown in Fig. 4, left). It can be seen that, despite the bell shape of the histograms, high ramp intensities seem to be related mostly to PC values in the ranges $[-20, -10]$ and $[0, 10]$. Thus, a certain non-linear dependence could be expected between these two variables since the knowledge of the PC at some time instant conveys some information about the probability of having a high ramp intensity at the same time instant.

4.2.1 Results for wind farm A: The results obtained for wind farm A are gathered in Fig. 7. This figure consists of four plots referring to the first four PCs considered for the MI assessment. Each plot displays eight stacks of bars according to the different cases described in (a) *variable selection*. Finally, each stack contains as many bars as spatial domains considered in (b) *domain selection*. The vertical axis indicates $MI_{\%}$, the MI as a percentage

$$\mathbf{X} = \begin{bmatrix} U_{10}^{1,(1 \times 1)} & V_{10}^{1,(1 \times 1)} & U_{10}^{1,(1 \times 2)} & V_{10}^{1,(1 \times 2)} & \dots & U_{10}^{1,(9 \times 9)} & V_{10}^{1,(9 \times 9)} \\ U_{10}^{2,(1 \times 1)} & V_{10}^{2,(1 \times 1)} & U_{10}^{2,(1 \times 2)} & V_{10}^{2,(1 \times 2)} & \dots & U_{10}^{2,(9 \times 9)} & V_{10}^{2,(9 \times 9)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ U_{10}^{N,(1 \times 1)} & V_{10}^{N,(1 \times 1)} & U_{10}^{N,(1 \times 2)} & V_{10}^{N,(1 \times 2)} & \dots & U_{10}^{N,(9 \times 9)} & V_{10}^{N,(9 \times 9)} \end{bmatrix}, \quad (7)$$

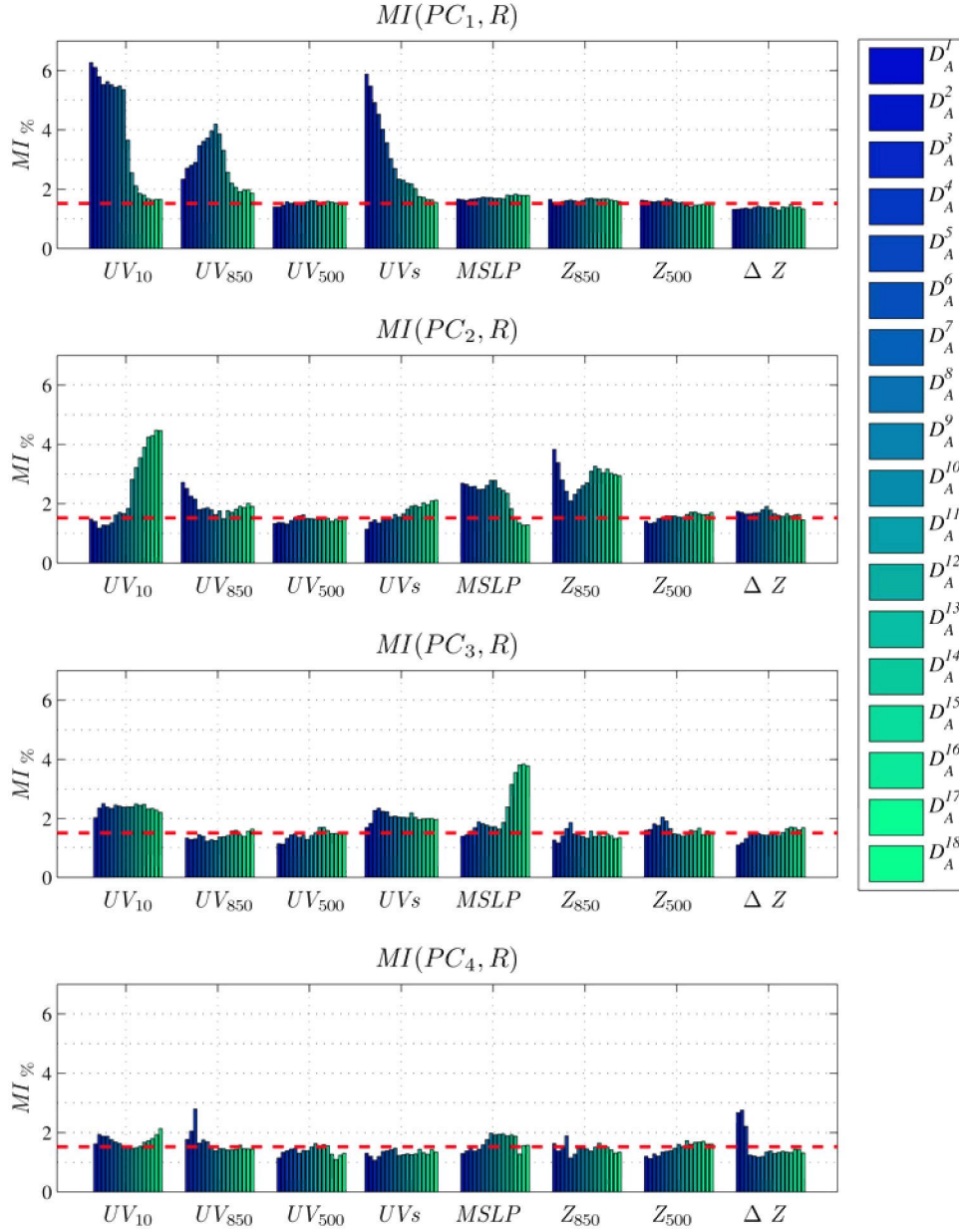


Fig. 7 $MI_{\%}$ between the ramp function at wind farm A and the PCs obtained by applying PCA to different combinations of meteorological variables and domains. The dashed line indicates the bias in estimating MI from a finite sample

of the entropy of the ramp function, $H(R)$. Thus, $MI_{\%} \in [0, 100]$. The horizontal dashed lines indicate the bias of the MI assessment.

Several conclusions can be drawn from Fig. 7. First, the MI estimates are, in general, very low. Indeed, they range from 1 to 6%, meaning that the uncertainty reduction of the ramp function derived from the knowledge of a certain PC is relatively limited. Furthermore, the lowest records fit with the estimated bias, indicating that these levels of MI are likely to be more a statistical effect due to the finite number of samples than a real dependence between the related variables.

Regarding the results related to wind fields [cases (a.1)–(a.4)], the highest levels of MI were attained for UV_{10} (in particular, by the two first PCs), outperforming the marks obtained for UV_{850} , which in turn overcame the results obtained for UV_{500} . In particular, the signal with the highest MI was found to be the first PC of the PCA applied to UV_{10} in the smaller domain considered, D_A^1 . It is possible to observe that larger domain sizes lead to a smooth decrease in $MI_{\%}$. However, it is worthy to mentioning that the abrupt decrease of MI observed for domains larger than D_A^9 has to do with the fact that the dominant pattern

(the channelling effect of the Ebro valley showed by \mathbf{eof}_1 in Fig. 4 left) becomes the second EOF in terms of explained variance; this also explains the abrupt increase in MI observed in the second PC. Conversely to UV_{10} , the PCs obtained for UV_{500} provided very low MI levels no matter the PC or the domain size considered, suggesting that ramp events at the wind farm level have little connection (if any) with the wind field at this pressure level. For the case of the PC_1 of the UV_{850} field, there seems to be an optimal domain size to seek relations between this meteorological variable and ramp occurrence; this domain size was found to be D_A^9 . However, this result does not necessarily mean that the wind field at this pressure level contributes with additional information in explaining ramp events as a visual inspection of the modes revealed that the related EOFs were found to merge features of UV_{10} and UV_{500} . In a similar manner, the high level of MI obtained for UV_s (first PC, smallest domain) may be due to the fact that the related EOF captures the same effect as the first mode of UV_{10} (i.e. the prevailing wind direction NW–SE determined by the channelling effect of the Ebro valley).

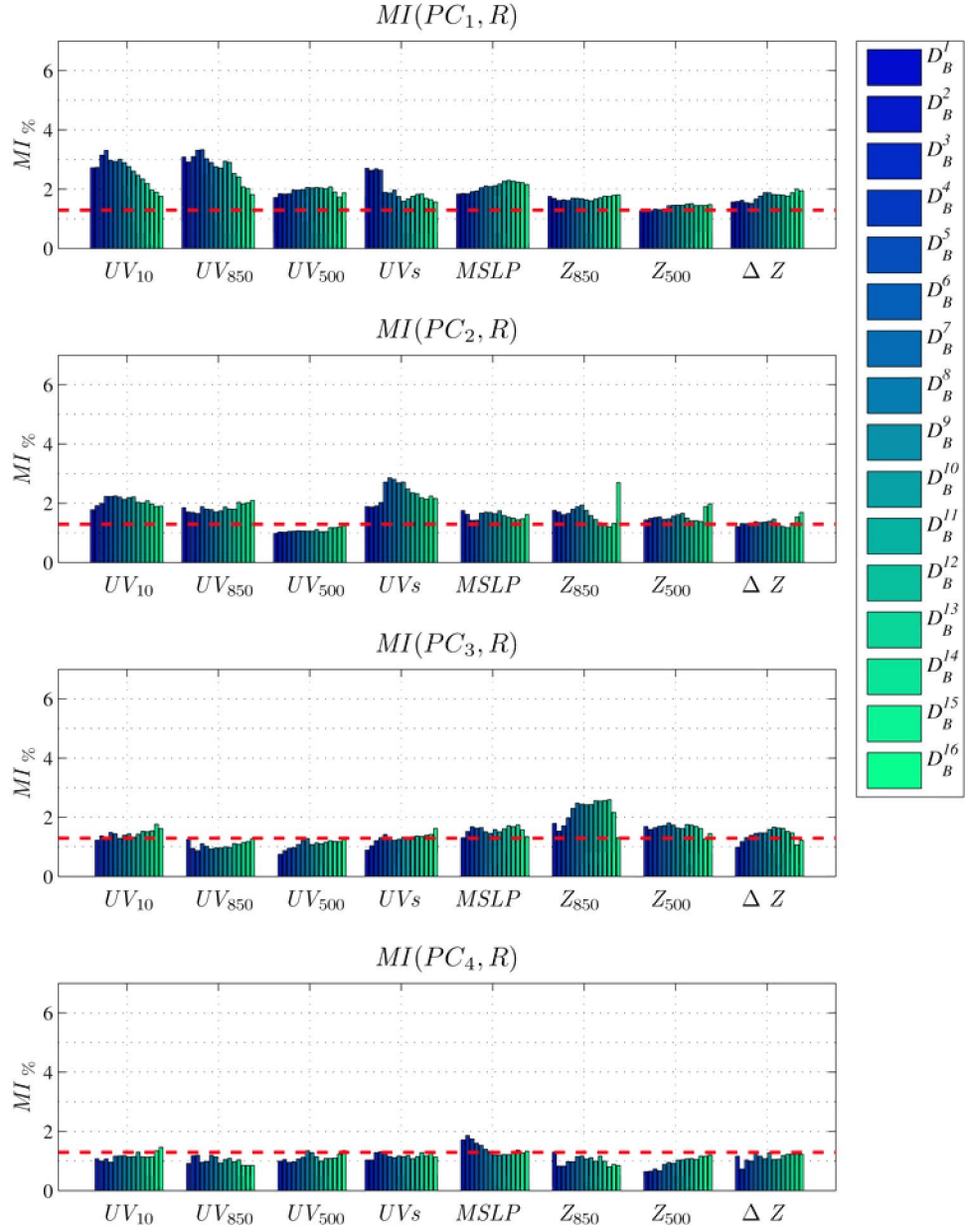


Fig. 8 $MI\%$ between the ramp function at wind farm B and the PCs obtained by applying PCA to different combinations of meteorological variables and domains. The dashed line indicates the bias in estimating MI from a finite sample

Results concerning scalar fields [cases (a.5)–(a.8)] were found to be, in general, of little less relevance, as they barely reached a $MI\%$ of 4%. For the case of the $MSLP$, it is noted that, despite the high level of explained variance of the first PC (see Fig. 5), the second and third PCs showed higher levels of MI than the first PC. This fact reveals that the presence of pressure gradients is more informative on ramp event occurrence than the average pressure level in the spatial domain, even when the latter accounts for most of the variance of this meteorological variable. Specifically, zonal pressure gradients detected on large spatial domains provided the highest MI scores among the scalar field features. A similar reasoning could be applied to the results obtained for the second PC of Z_{850} , which is related to meridional gradients. On the other hand, cases involving meteorological variables at high altitude (Z_{500} and ΔZ) did not show either relevant MI levels or any kind of dependence with the PC or the domain size considered.

4.2.2 Results for wind farm B: Fig. 8 displays the MI between the ramp function at wind farm B and the set of PCs obtained from the PCAs. The vertical axis was scaled as in Fig. 7 so as to point out a significant result: that the uncertainty reduction of the

ramp occurrence derived from the knowledge of the PCs were found to be notably lower in this case study. It can be observed that the levels of MI barely exceeded 3%, half of the results

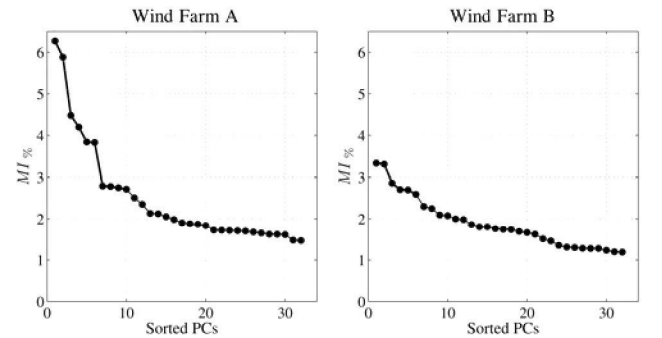


Fig. 9 MI marks sorted in decreasing order (see Section 4.3 for details)

attained at wind farm A. In addition, considering different domain sizes or meteorological variables revealed little impact on the MI estimates, which were in many cases similar to the estimated bias (dashed line).

Concerning wind fields, results for UV_{10} and UV_{850} were qualitatively similar. Unlike what was observed at wind farm A, second and third PCs of $MSLP$ provided lower MI levels than the first PC, suggesting that pressure gradients are likely to be less informative on ramp events than the average pressure level. The best results among the scalar fields were obtained for the second PC (domain D_B^{16}) and third PC (domain D_B^{14}) of Z_{850} . In both cases, the associated EOF was related to a zonal gradient of the field. Finally, results concerning Z_{500} and ΔZ showed very low MI marks, which is in line with those results obtained for wind farm A.

4.3 Further discussion

The analysis presented above shows how specific features relevant in explaining wind power ramp occurrence can be identified in a large dataset. It is reminded that the objective here is to facilitate the definition of low-dimensional signals meaningful for ramp forecasting. Concerning the results obtained for wind farm A, it was found that PCs related to regional weather regimes and zonal pressure gradients could represent a good choice to feed a ramp forecasting model as compared with other candidates derived from the same dataset. The link between pressure gradients and ramp occurrence has been identified in previous works ([26] in Tasmania and [27] in Portugal, among others). Thus, the proposed methodology seemed to work well in identifying relevant spatial domains, meteorological variables and meteorological features. Interestingly, results concerning wind farm B suggested a weaker link of ramp events with the global scale. To better reflect this idea, the highest mark of each stack of bars in Figs. 7 and 8 has been retained and sorted in decreasing order (see Fig. 9). This fact might be due to either the inherent limitations of the proposed methodology or the fact that a limited dataset was employed (reduced the number of meteorological variables considered and spatial domain restricted to the Iberian Peninsula). However, the hypothesis of different degrees of connection between the global scale and the wind power dynamics at the two considered case studies is also postulated. Indeed, ramps can be motivated by meteorological processes occurring at several spatial/temporal scales. Thus, the use of GCM outputs limits the analysis to those events caused by meteorological processes resolved by the atmospheric model.

At this point, in order to put the results of the analysis in context, a description of potential limitations of the methodology is provided. An important limitation of the methodology deals with the use of PCA to reveal features in the data. Indeed, EOFs represent the way in which the variables involved in the analysis combine to jointly exhibit maximum variance. For the case of meteorological variables, these modes were interpreted as relevant atmospheric patterns in an effort to identify underlying physical mechanisms that dominate the dynamics in the considered domain. However, it is important to bear in mind that the EOFs result from the data, but also from the statistical tool employed, which imposes specific conditions to the results obtained. In particular, PCA provides orthogonal modes and uncorrelated PCs. This may come into conflict when relating EOFs with physical processes because the main modes of variability of a non-linear and highly complex system as the atmosphere are not necessarily orthogonal or uncorrelated [28, 29]. In order to address this issue, the technique known as *rotated PCA* could be employed, though its application requires some arbitrary decisions that also condition to some extent the patterns obtained [30].

Probably, the main limitation of the proposed approach has to do with the ramp characterisation. Indeed, the employed ramp function only conveys observed ramps, that is, ramps that *finally happened*. However, observed ramps merely represent a portion of the situations that ramp forecasting should address, and the ramp function fails in capturing situations where wind power volatility

is high even when a ramp event did not happened. As an example, let us think of a wind farm operating at a wind speed very close to the cut-out wind speed. In such situation, the probability of having a ramp-down event is higher than operating at nominal wind speed, but the ramp function is unable to capture the associated risk because it is built from the observed power output. This fact is likely to affect specially TSOs because assessing backup reserves requires dealing with uncertainty. Thus, a proper MI assessment (Section 4.2) would require the ramp function to be upgraded so that it conveys also the probability of ramp. This issue would probably require the use of advanced techniques dealing with generation scenarios and probabilistic forecast, which is out of the scope of this work, though represents a clear path for improvement.

5 Conclusions

To improve ramp forecasting, ramp underlying causes need to be explored. This requires dealing with large datasets derived from monitoring/modelling the stages involved in the wind-to-power conversion chain. To this end, a methodological proposal oriented to systematically relate ramp events to massive datasets was presented.

As a particular case, the proposed methodology was employed to explore the relationships between atmospheric dynamics at the global/synoptic scales and ramp events experienced at the wind farm level. To this end, reanalysis data generated with a GCM were employed together with generation data originated from two wind farms located in Spain. This analysis has provided a preliminary quantification on what can be expected from this method. In particular, experimental results allowed us to identify spatial domains and atmospheric circulation features that are to a certain degree relevant in explaining ramp occurrence for one of the wind farms. Specifically, the link between the reanalysis dataset and ramp events at this wind farm was found to be mainly associated to regional weather regimes governed by the channelling effect of the Ebro valley (Cierzo and Bochorio) and the presence of zonal pressure gradients over the Iberian Peninsula, which are likely to be indicative of the passage of large scale weather systems. Results concerning the second wind farm suggested a weaker link of ramp events with the global scale.

An important point of the proposed methodology is that it can be readily applied to datasets related to other stages of the wind-to-power conversion process. This would eventually provide a comprehensive understanding of ramp underlying causes. In addition, we point out that the outputs of the analysis could eventually allow for the elaboration of meaningful explanatory variables to feed wind power ramp forecasting models. This is so because the methodology extracts low-dimensional information from big data on the basis of the estimated dependence with ramp events.

6 Acknowledgment

EGB was funded by the UMU (University of Murcia) - Incoming Mobility Programme ACTION - COFUND- Marie Curie Actions.

7 References

- Costa, A., Crespo, A., Navarro, J., *et al.*: 'A review on the young history of the wind power short-term prediction', *Renew. Sustain. Energy Rev.*, 2008, **12**, (6), pp. 1725–1744
- Giebel, G.: 'The state of the art in short-term prediction of wind power – a literature overview (2nd ed.)'. Technical report, ANEMOS.plus/SafeWind projects, 2011
- Ferreira, C., Gama, J., Matias, L., *et al.*: 'A survey on wind power ramp forecasting'. Technical report, Argonne National Laboratory, 2010. Available at <http://www.dis.anl.gov/pubs/69166.pdf>
- Potter, C.W., Grimit, E., Nijssen, B.: 'Potential benefits of a dedicated probabilistic rapid ramp event forecast tool'. IEEE/PES Power Systems Conf. and Exposition, Seattle, Washington, March 2009, pp. 1–5

- 5 Zack, J.W.: 'Optimization of wind power production forecast performance during critical periods for grid management'. Technical report, June 2007. Available at (March 23, 2011): http://www.awstruepower.com/wp-content/media/2010/05/AWEA_Windpower_2007_Forecasting.pdf.pdf
- 6 Cutler, N., Kay, M., Jacka, K., *et al.*: 'Detecting, categorizing and forecasting large ramps in wind farm power output using meteorological observations and WPPT', *Wind Energy*, 2007, **10**, (5), pp. 453–470
- 7 AWS Truewind: 'AWS Truewind's final report for the Alberta forecasting pilot project'. Technical report, 2008. Available at (March 23, 2011): http://www.uwig.org/Alberta_PP_Final_Report_AWST_Jun25.pdf
- 8 Bossavy, A., Girard, R., Kariniotakis, G.: 'Forecasting ramps of wind power production with numerical weather prediction ensembles', *Wind Energy*, 2013, **16**, (1), pp. 51–63
- 9 Kamath, C.: 'Associating weather conditions with ramp events in wind power generation'. IEEE/PES Power Systems Conf. and Exposition, Phoenix, Arizona, March 2011, pp. 1–8
- 10 Kamath, C.: 'Understanding wind ramp events through analysis of historical data'. IEEE PES Transmission and Distribution Conf. and Exposition: Smart Solutions for a Changing World, New Orleans, LA, April 2010
- 11 Gallego, C., Costa, A., Cuerva, A.: 'Improving short-term forecasting during ramp events by means of regime-switching artificial neural networks', *Adv. Sci. Res.*, 2011, **6**, pp. 55–58
- 12 Gallego, C., Costa, A., Cuerva, A., *et al.*: 'A wavelet-based approach for large wind power ramp characterisation', *Wind Energy*, 2013, **16**, (2), pp. 257–278
- 13 Wilks, D.S.: 'Statistical methods in the atmospheric sciences' (Elsevier Academic Press Publications, Burlington, USA, 2006, 2nd edn.)
- 14 Dionisio, A., Menezes, R., Mendes, D.A.: 'Mutual information: a measure of dependency for nonlinear time series', *Phys. A Stat. Mech. Appl.*, 2004, **344**, (1–2), pp. 326–329
- 15 Steuer, R., Kurths, J., Daub, C.O., *et al.*: 'The mutual information: Detecting and evaluating dependencies between variables', *Bioinformatics*, 2002, **18**(suppl 2), pp. S23–S240
- 16 Shannon, C.E.: 'A mathematical theory of communication', *Bell Syst. Techn. J.*, 1948, **27**, pp. 379–423, 623–656
- 17 Grassberger, P.: 'Finite sample corrections to entropy and dimension estimates', *Phys. Lett. A*, 1988, **128**, (6–7), pp. 369–373
- 18 Gallego, C.: 'Statistical models for short-term wind power ramp forecasting'. PhD thesis, Escuela Técnica Superior de Ingenieros Aeronáuticos, Universidad Politécnica de Madrid, 2013. Available at: <http://oa.upm.es/21912>
- 19 Saha, S., Moorthi, S., Pan, H.L. *et al.*: 'The NCEP climate forecast system reanalysis', *Bull. Am. Meteorol. Soc.*, 2010, **91**, (8), pp. 1015–1057
- 20 Reanalysis Intercomparison and Observations. <http://www.reanalyses.org/atmosphere/comparison-table>. Accessed date: July 20, 2014
- 21 García-Bustamante, E., González-Rouco, J.F., Navarro, J., *et al.*: 'North Atlantic atmospheric circulation and surface wind in the Northeast of the Iberian Peninsula: uncertainty and long term downscaled variability', *Clim. Dyn.*, 2012, **38**, (1–2), pp. 141–160
- 22 Jiménez, P.A., González-Rouco, J.F., Montávez, J.P., *et al.*: 'Climatology of wind patterns in the northeast of the Iberian Peninsula', *Int. J. Climatol.*, 2009, **29**, (4), pp. 501–525
- 23 Lange, M., Focken, U.: 'Physical approach to short-term wind power prediction' (Springer, Berlin, 2005)
- 24 García-Bustamante, E., González-Rouco, J.F., Navarro, J., *et al.*: 'Relationship between wind power production and North Atlantic atmospheric circulation over the northeastern Iberian Peninsula', *Clim. Dyn.*, 2013, **40**, (3–4), pp. 935–949
- 25 Wallace, J.M., Hobbs, P.V.: 'Atmospheric science: an introductory survey' (International Geophysics Series. Academic Press, 2006, 2nd edn.)
- 26 Cutler, N.J., Outhred, H.R., MacGill, I.F., *et al.*: 'Characterizing future large, rapid changes in aggregated wind power using numerical weather prediction spatial fields', *Wind Energy*, 2009, **12**, (6), pp. 542–555
- 27 Couto, A., Costa, P., Rodrigues, L., *et al.*: 'Impact of weather regimes on the wind power ramp forecast in Portugal', *IEEE Trans. Sustain. Energy*, 2014, **PP**, (99), pp. 1–9
- 28 Monahan, A.H., Fyfe, J.C., Ambaum, M.H.P., *et al.*: 'Empirical orthogonal functions: the medium is the message', *J. Clim.*, 2009, **22**, pp. 6501–6514
- 29 North, G.R.: 'Empirical orthogonal functions and normal modes', *J. Atmos. Sci.*, 1984, **41**, (5), pp. 879–887
- 30 Hannachi, A., Jolliffe, I.T., Stephenson, D.B.: 'Empirical orthogonal functions and related techniques in atmospheric science: a review', *Int. J. Climatol.*, 2007, **27**, (9), pp. 1119–1152